



**PS204 Introduction à l'apprentissage statistique**

**Semestre 8**

---

## Rapport du Projet

---

**Télécommunication - ENSEIRB-MATMECA**

*Réalisé par :*

MOHAMED AHALLI  
HICHAM FELLAH  
IMAD ABIED

*Encadré par :*

P. VALLET

---

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Principes de l'ACP</b>	<b>3</b>
<b>3</b>	<b>Réduction de dimension et eigenfaces</b>	<b>7</b>
3.1	Q2 : Représentation des eigenfaces . . . . .	7
3.2	Q3 : l'évolution de l'image reconstruite après ACP . . . . .	7
3.3	Q4 : Ratio de reconstruction . . . . .	9
3.4	Q5 : Influence de l'ensemble d'entraînement : . . . . .	10
<b>4</b>	<b>Classification :</b>	<b>10</b>
4.1	Q1 : Implémentation du classifieur K-NN . . . . .	10
4.2	Q2 : Calculer les matrices de confusion sur les bases de test . . . . .	11
4.3	Q3 : Représentation des nuages de points . . . . .	13
4.4	Q4 : Implémentation du classifieur guassien . . . . .	13
4.5	Q5 : les expérimentations menées pour le classifieur guassien . . . . .	14
4.6	Q6 : Modification du classifieur gaussien . . . . .	16

## Table des figures

1	n eigenfaces . . . . .	7
2	Reconstruction pour $l = 8$ ( $\alpha = 0.9$ ) . . . . .	8
3	Reconstruction pour $l = 3$ ( $\alpha = 0.5$ ) . . . . .	8
4	Reconstruction pour $l = 1$ ( $\alpha = 0.2$ ) . . . . .	9
5	Ratio de reconstruction . . . . .	10
6	Test de 1 à 3 . . . . .	11
7	Test de 4 à 6 . . . . .	12
8	les nuages de points des couples de composantes principales (1, 2),(2, 3),(3, 4),(4, 5) . . . . .	13
9	Test de 1 à 2 . . . . .	14
10	Test de 3 à 4 . . . . .	15
11	Test de 5 à 6 . . . . .	16

---

## 1 Introduction

La théorie de l'apprentissage statistique traite du problème de la recherche d'une fonction prédictive basée sur des données. La théorie de l'apprentissage statistique a conduit à des applications dans des domaines tels que la vision par ordinateur, la reconnaissance de la parole, la bioinformatique. Les objectifs de l'apprentissage sont la prédiction et la compréhension. L'apprentissage relève de plusieurs catégories, y compris l'apprentissage supervisé, l'apprentissage non supervisé, l'apprentissage en ligne et l'apprentissage par renforcement.

## 2 Principes de l'ACP

**L'analyse en composantes principales**, et plus généralement de la statistique multivariée, qui consiste à transformer des variables liées entre elles (dites « corrélées » en statistique) en nouvelles variables décorrélées les unes des autres. Ces nouvelles variables sont nommées « composantes principales » ou axes principaux. Elle permet au statisticien de résumer l'information en réduisant le nombre de variables.

Il s'agit d'une approche à la fois géométrique (les variables étant représentées dans un nouvel espace, selon des directions d'inertie maximale) et statistique (la recherche portant sur des axes indépendants expliquant au mieux la variabilité — la variance — des données) .

Considérons un ensemble de données  $x_1, \dots, x_n$  de  $R^p$ , dite données d'apprentissage, dont on note :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

La moyenne empirique est :

$$\hat{R} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T$$

on cherche une base de vecteurs orthonormés  $v_1, \dots, v_l R^p$ , tels que les variances empiriques :

$$E_j = \frac{1}{n} \sum_{i=1}^n \langle v_j, x_i - \bar{x} \rangle^2, j = 1, \dots, l$$

1. soit  $v \in \mathbb{R}^p$

$$\begin{aligned} v^\top \hat{R} v &= \frac{1}{n} \sum_{i=1}^n v(x_i - \bar{x})(x_i - \bar{x})^\top v^\top \\ &= \frac{1}{n} \sum_{i=1}^n v(x_i - \bar{x})(v(x_i - \bar{x}))^T \\ &= \frac{1}{n} \sum_{i=1}^n \langle v, x_i - \bar{x} \rangle^2 \end{aligned}$$

$$\sup_{\substack{v \in \mathbb{R}^p \\ \|v\|_2=1}} \frac{1}{n} \sum_{i=1}^n \langle v, x_i - \bar{x} \rangle^2 = \sup_{\substack{v \in \mathbb{R}^p \\ \|v\|_2=1}} (v^\top \hat{R} v) = \rho(\hat{R}) = \lambda_1 = \|\hat{R}\|$$

où  $\rho$  est le rayon spectral de  $\hat{R}$ , la dernière égalité est justifiée par  $\hat{R}$  est symétrique.

Donc :

$$u_1 \in \underset{\substack{v \in \mathbb{R}^d, \|v\|_2=1}}{\operatorname{argmax}} \left( \frac{1}{n} \sum_{i=1}^n \langle v, x_i - \bar{x} \rangle^2 \right)$$

D'où

$$E_1 = \lambda_1$$

2. soit  $V \in \mathbb{R}^d$

Pour  $j = 2, v \perp u_1, \|v\|_2 = 1$

On décompose notre espace de travaille  $\mathbb{R}^p$  comme étant la somme orthogonale  $\text{Vect}(u_1)^\perp \oplus E$  où  $E = \text{Vect}(u_2, \dots, u_p)$  l'otogonal de  $\text{Vect}(u_1)$

Soit  $\hat{R}_E$  la matrice induite de l'espace  $E$

Alors  $\hat{R}_E = \text{diag}(\lambda_2, \dots, \lambda_p)$ , on applique ce qui précède sur cette nouvelle matrice on aura :

$$\sup_{\substack{v \in \mathbb{R}^d, \|v\|_2=1 \\ v \perp u_1}} \frac{1}{n} \sum_{i=1}^n \langle v, x_i - \bar{x} \rangle^2 = \lambda_2$$

D'où

$$u_2 \in \underset{\substack{v \in \mathbb{R}^d, \|v\|_2=1 \\ v \perp u_1}}{\operatorname{argmax}} \left( \frac{1}{n} \sum_{i=1}^n \langle v, x_i - \bar{x} \rangle^2 \right)$$

Ainsi de la même manière en décomposant  $\mathbb{R}^p$  comme deux espaces orthogonaux on trouve :

D'où  $\forall j > 1$

$$\begin{aligned} u_j &\in \underset{\substack{v \in \mathbb{R}^d, \|v\|_2=1 \\ v \perp u_1, \dots, u_{j-1}}}{\operatorname{argmax}} \left( \frac{1}{n} \sum_{i=1}^n \langle v, x_i - \bar{x} \rangle^2 \right) \\ E_j &= \frac{1}{n} \sum_{i=1}^n \langle v_j, x_i - \bar{x} \rangle^2 = \lambda_j \end{aligned}$$

3. On a

$$\begin{aligned}
 \frac{1}{n} \sum_{i=1}^n \|\pi_s(x_1 - \bar{x})\|_2^2 &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^l \langle u_j, x_i - \bar{x} \rangle^2 \\
 &= \sum_{j=1}^l \left( \frac{1}{n} \sum_{i=1}^n \langle u_j, x_i - \bar{x} \rangle^2 \right) \\
 &= \sum_{j=1}^l \lambda_j
 \end{aligned}$$
  

$$\begin{aligned}
 \frac{1}{n} \sum_{i=1}^n \|x_i - \bar{x}\|_2^2 &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p \langle u_j, x_i - \bar{x} \rangle^2 \\
 &= \sum_{j=1}^p \left( \frac{1}{n} \sum_{i=1}^n \langle u_j, x_i - \bar{x} \rangle^2 \right) \\
 &= \sum_{j=1}^p \lambda_j
 \end{aligned}$$

La permutation des sommes est justifiée par le fait que les sommes sont finies.

D'où : r

$$k(l) = \frac{\frac{1}{n} \sum_{i=1}^n \|\pi_s(x_i - \bar{x})\|_2^2}{\frac{1}{n} \sum_{i=1}^n \|x_i - \bar{x}\|_2^2} = \frac{\sum_{j=1}^l \lambda_j}{\sum_{j=1}^p \lambda_j}$$

4. On a :

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \underline{1}^T \underline{X}, \text{ Où } \underline{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \text{ et } \underline{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Donc

$$\begin{aligned}
 \underline{X}^T &= \frac{1}{\sqrt{n}} [x_1 - \bar{X}, \dots, x_n - \bar{X}] \\
 &= \frac{1}{\sqrt{n}} (\underline{X} - \bar{X} \underline{1}) \\
 &= \frac{1}{\sqrt{n}} (\underline{X} - \frac{1}{n} \underline{1}^T \underline{1} \underline{X}) \\
 &= \frac{1}{\sqrt{n}} (I - \frac{1}{n} \underline{1}^T \underline{1}) \underline{X} \\
 &= \underline{\pi} \underline{X}
 \end{aligned}$$

Où  $\underline{\pi} = \frac{1}{\sqrt{n}} (I - \frac{1}{n} \underline{1}^T \underline{1})$  la projection sur un espace de dimension n-1

Donc :

$$\text{rg}(\underline{X}) = n-1$$

5. Si  $v$  est vecteur propre de  $X^T X$  Donc  $\exists \lambda$  tq  $X^T X v = \lambda v$

On a :  $XX^T X v = X \lambda v = \lambda X v = \lambda u$

Donc :

$u = Xv$  est un vecteur propre de  $XX^T$

6. Soient  $v_1, \dots, v_{n-1}$  un ensemble de vecteur propre orthonormée de  $X^T X$

On a

$$XV = X[v_1, \dots, v_{n-1}] = [Xv_1, \dots, Xv_{n-1}] = [u_1, \dots, u_{n-1}]$$

Donc :

$U = XV[V^T X^T X V]^{-1/2}$

est une matrice à colonnes orthonormées de vecteurs propres de  $\hat{R}$  associées aux valeurs propres non nuls.

N.B : le terme  $[V^T X^T X V]^{-1/2}$  est une valeur qui peut s'écrire sous la forme  $\sum_1^n \lambda_i u_i^2$

### 3 Réduction de dimension et eigenfaces

#### 3.1 Q2 : Représentation des eigenfaces

Les eigenfaces constituent une base orthonormée décrivant les visages de la base d'apprentissage.



FIGURE 1 – n eigenfaces

#### 3.2 Q3 : l'évolution de l'image reconstruite après ACP

Nous observons que le choix de  $l$  a un impact direct sur les traits des visages ; plus le  $l$  est petit, plus les visages reconstruits se rassemblent. Et plus on augmente  $l$  plus les nuances entre les visages s'élargissent.



FIGURE 2 – Reconstruction pour  $l = 8$  ( $\alpha = 0.9$ )



FIGURE 3 – Reconstruction pour  $l = 3$  ( $\alpha = 0.5$ )



FIGURE 4 – Reconstruction pour  $l = 1$  ( $\alpha = 0.2$ )

### 3.3 Q4 : Ratio de reconstruction

Nous constatons que le facteur de qualité saute très rapidement et dépasse le 0.9 à partir de  $l=8$ . Cela nous rassure à l'égard des conséquences de la réduction de 1 citées dans la question précédente.

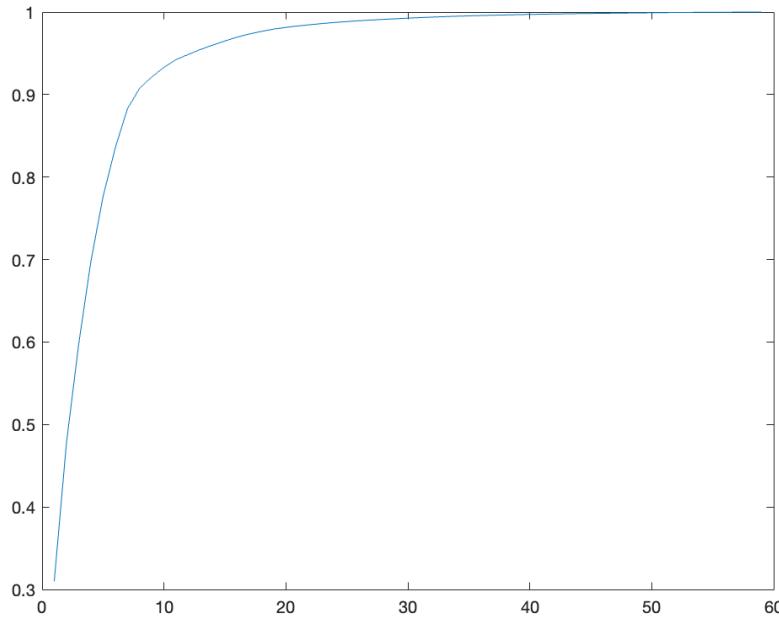


FIGURE 5 – Ratio de reconstruction

### 3.4 Q5 : Influence de l'ensemble d'entraînement :

Le data set training 2 contient deux fois moins de photos de visages par rapport au data set training 1. Par conséquent la matrice de covariance empirique sera moins significante. Par la suite, il se peut que les premiers vecteurs propres choisis par l'ordre de leurs valeurs propres soient pas les vecteurs qui comptent le plus pour une image qui n'appartient pas à l'ensemble d'entraînement.

## 4 Classification :

### 4.1 Q1 : Implémentation du classifieur K-NN

#### Compromis lors du choix de k :

Si  $k$  est grande (égale au nombre total des images par exemple), l'algorithme de classifieur k-NN va perdre son sens parce que, dans ce cas, on va choisir la classe la plus présente dans l'ensemble d'entraînement et qui ne dépend même pas à l'image à classifier.

En revanche, si on choisit un  $k$  petit (égal à 1 par exemple), la classification va se reposer seulement sur l'image de l'ensemble d'entraînement qui minimise la distance

---

## 4 CLASSIFICATION :

euclidienne par rapport à l'image à classifier. Cela ne signifie pas forcément que l'image trouvée a la même classe que l'image à classifier mais il peut signifier aussi que les deux images étaient prises dans les mêmes conditions de luminosité par exemple.

## 4.2 Q2 : Calculer les matrices de confusion sur les bases de test

```

Test 1 :
Confusion matrix :
C =

1      0      0      0      0      0
0      1      0      0      0      0
0      0      1      0      0      0
0      0      0      1      0      0
0      0      0      0      1      0
0      0      0      0      0      1

error rate : 0.000000
-----

Test 2 :
Confusion matrix :
C =

1      0      0      0      0      0
0      1      0      0      0      0
0      0      1      0      0      0
0      0      0      1      0      0
0      0      0      0      1      0
0      0      0      0      0      1

error rate : 0.000000
-----
```

Test 3 :

```

Confusion matrix :
C =
```

0.6364	0	0.1818	0	0	0.1818
0	0.9091	0	0.0909	0	0
0	0	1.0000	0	0	0
0	0	0	1.0000	0	0
0	0	0.1818	0	0.6364	0.1818
0	0	0.0909	0.0909	0	0.8182

```

error rate : 0.166667
```

FIGURE 6 – Test de 1 à 3

```
Test 4 :
Confusion matrix :
C =
```

0.3333	0	0.6667	0	0	0
0	0.5556	0.3333	0.1111	0	0
0	0	0.8889	0.1111	0	0
0	0	0	1.0000	0	0
0	0	0.5556	0.1111	0.2222	0.1111
0	0	0.2222	0.3333	0	0.4444

```
error rate : 0.425926
```

```
-----
```

```
Test 5 :
Confusion matrix :
C =
```

0	0	0.7333	0.2667	0	0
0	0	0.4000	0.6000	0	0
0	0	0.9333	0.0667	0	0
0	0	0.2667	0.7333	0	0
0	0	0.8000	0.2000	0	0
0	0	0.1333	0.8000	0	0.0667

```
error rate : 0.711111
```

```
-----
```

```
Test 6 :
Confusion matrix :
C =
```

0	0	0	1.0000	0	0
0	0	0	1.0000	0	0
0	0	0.2500	0.7500	0	0
0	0	0.0833	0.9167	0	0
0	0	0.2500	0.7500	0	0
0	0	0	1.0000	0	0

```
error rate : 0.805556
```

FIGURE 7 – Test de 4 à 6

On remarque que le taux de mauvaise classification augmente de plus en plus du test 1 au test 6, cela dû au fait que la matrice de confusion est de plus en plus moins proche de la matrice identité. Du coup la classification est moins performante, et c'est normal car la luminosité des images du test 6 diminue dans plusieurs parties du visage ce qui rend difficile la classification.

### 4.3 Q3 : Représentation des nuages de points

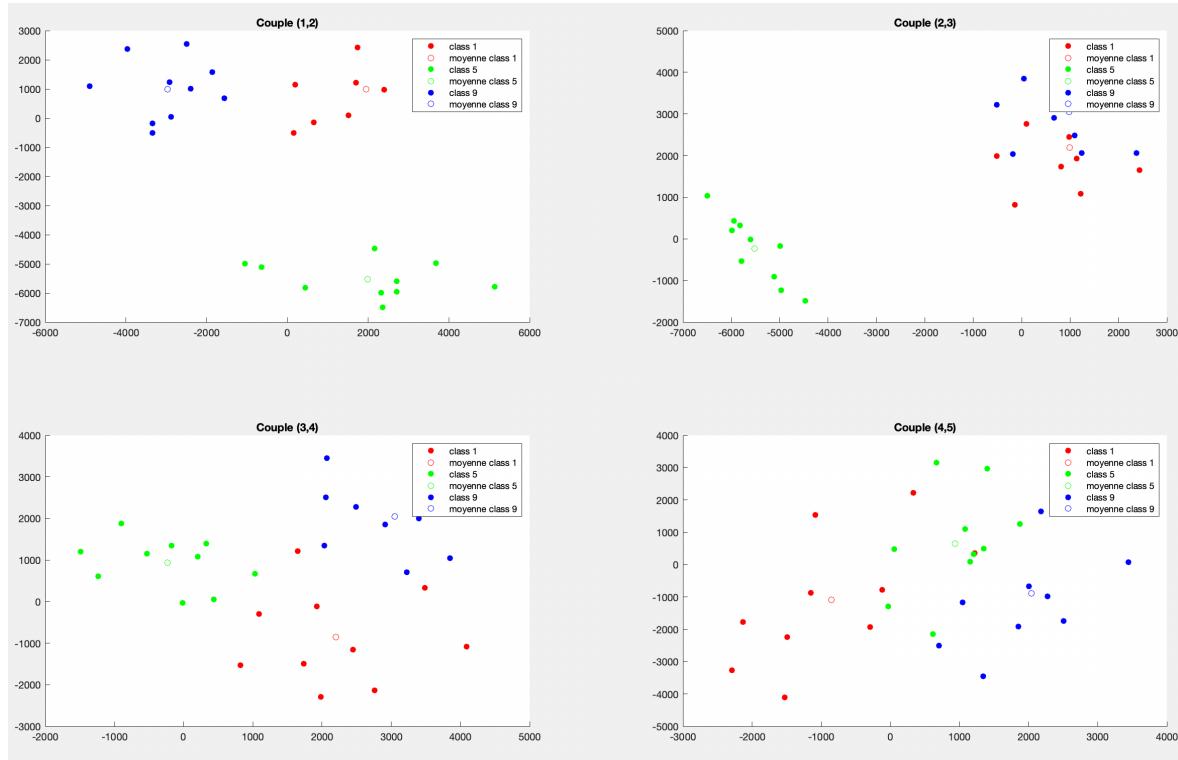


FIGURE 8 – les nuages de points des couples de composantes principales (1, 2),(2, 3),(3, 4),(4, 5)

On remarque que les moyennes intra-classe sont situées aux centres des nuages des points. Nous constatons aussi que le nuage des points de chaque catégorie de visage est dense autour de la moyenne. Puis la densité commence à diminuer dès qu'on s'éloigne de la moyenne. Cela est vrai pour le couple (1, 2), (2, 3), (3, 4), (4, 5). Si on suppose que cela reste vrai pour les couples restants, alors on peut dire que le choix d'un modèle gaussien a été judicieux.

### 4.4 Q4 : Implémentation du classifieur guassien

#### Limitation potentiel de classifieur de gauss :

L'implémentation de classifieur gaussien proposée par le sujet ne prend pas en compte le cas où les matrices de covariance sont différentes d'une catégorie de visage à une autre. Par conséquent, si une catégorie de visage se caractérise par une matrice de covariance très différente par rapport aux autres, alors dans ce cas, soit sa fonction de densité va maximiser le critère de vraisemblance alors qu'elle ne doit pas, soit il sera incapable de maximiser le critère de vraisemblance alors qu'il censé le maximiser.

## 4.5 Q5 : les expérimentations menées pour le classifieur guassien

```
-----  
Test 1 :  
Confusion matrix :  
C =  
  
1 0 0 0 0 0  
0 1 0 0 0 0  
0 0 1 0 0 0  
0 0 0 1 0 0  
0 0 0 0 1 0  
0 0 0 0 0 1  
  
error rate : 0.000000  
  
-----  
Test 2 :  
Confusion matrix :  
C =  
  
1 0 0 0 0 0  
0 1 0 0 0 0  
0 0 1 0 0 0|  
0 0 0 1 0 0  
0 0 0 0 1 0  
0 0 0 0 0 1  
  
error rate : 0.000000  
  
-----
```

FIGURE 9 – Test de 1 à 2

Test 3 :

Confusion matrix :

C =

1	0	0	0	0	0
0	1	0	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0
0	0	0	0	1	0
0	0	0	0	0	1

error rate : 0.000000

---

Test 4 :

Confusion matrix :

C =

1.0000	0	0	0	0	0
0	1.0000	0	0	0	0
0	0	1.0000	0	0	0
0	0	0.2222	0.7778	0	0
0	0	0	0	1.0000	0
0	0	0.3333	0	0	0.6667

error rate : 0.092593

---

FIGURE 10 – Test de 3 à 4

```
-----
Test 5 :
Confusion matrix :
C =

    0.2667      0     0.7333      0      0      0
        0      0     1.0000      0      0      0
        0      0     1.0000      0      0      0
        0      0     1.0000      0      0      0
        0      0     0.9333      0   0.0667      0
        0      0     0.6667      0      0   0.3333

error rate : 0.722222

-----
Test 6 :
Confusion matrix :
C =

        0      0     1.0000      0      0      0
        0      0     1.0000      0      0      0
        0      0     1.0000      0      0      0
    0.0833      0     0.9167      0      0      0
        0      0     1.0000      0      0      0
    0.0833      0     0.9167      0      0      0

error rate : 0.833333
```

FIGURE 11 – Test de 5 à 6

En comparant avec les tests pour le classifieur k-NN, on remarque que le classifieur gaussien est plus performant, notamment pour les test 3 et 4. Par contre pour les tests 5 et 6 les résultats sont similaires.

#### 4.6 Q6 : Modification du classifieur gaussien

Nous proposons de calculer une matrice de covariance pour chaque catégorie de visage. Dans ce sens, nous donnons l'expression de la matrice de covariance ainsi :

$$\hat{\Sigma}_j = \frac{1}{|\mathbb{C}_j|} \sum_{i \in \mathbb{C}_j} (\omega(x_i) - \hat{\mu}_j)(\omega(x_i) - \hat{\mu}_j)^T$$

L'avantage de cette approche c'est qu'elle permet d'avoir une classification plus précise et permet aussi de pallier la limitation citée dans la question de l'implémentation de classifieur de gauss.

L'inconvénient de cette approche c'est la complexité calculatoire de  $\varphi(x)$  qui ne sera plus simplifiable comme précédent car la matrice de covariance dépend des classes.